

On prefixal factorizations of words

Aldo de Luca^a, Luca Q. Zamboni^{c,b}

^a*Dipartimento di Matematica e Applicazioni, Università di Napoli Federico II, Italy*

^b*FUNDIM, University of Turku, Finland*

^c*Université de Lyon, Université Lyon 1, CNRS UMR 5208, Institut Camille Jordan, 43 boulevard du 11 novembre 1918, F69622 Villeurbanne Cedex, France*

Abstract

We consider the class \mathcal{P}_1 of all infinite words $x \in \mathbb{A}^\omega$ over a finite alphabet \mathbb{A} admitting a prefixal factorization, i.e., a factorization $x = U_0U_1U_2 \cdots$ where each U_i is a non-empty prefix of x . With each $x \in \mathcal{P}_1$ one naturally associates a “derived” infinite word $\delta(x)$ which may or may not admit a prefixal factorization. We are interested in the class \mathcal{P}_∞ of all words x of \mathcal{P}_1 such that $\delta^n(x) \in \mathcal{P}_1$ for all $n \geq 1$. Our primary motivation for studying the class \mathcal{P}_∞ stems from its connection to a coloring problem on infinite words independently posed by T. Brown in [3] and by the second author in [17]. More precisely, let \mathbf{P} be the class of all words $x \in \mathbb{A}^\omega$ such that for every finite coloring $\varphi : \mathbb{A}^+ \rightarrow C$ there exist $c \in C$ and a factorization $x = V_0V_1V_2 \cdots$ with $\varphi(V_i) = c$ for each $i \geq 0$. In [5] we conjectured that a word $x \in \mathbf{P}$ if and only if x is purely periodic. In this paper we show that $\mathbf{P} \subseteq \mathcal{P}_\infty$, so in other words, potential candidates to a counter-example to our conjecture are amongst the non-periodic elements of \mathcal{P}_∞ . We establish several results on the class \mathcal{P}_∞ . In particular, we show that a Sturmian word x belongs to \mathcal{P}_∞ if and only if x is nonsingular, i.e., no proper suffix of x is a standard Sturmian word.

Keywords: Combinatorics on words, Prefixal factorization, Sturmian word, Coloring problems

2010 MSC: 68R15

1. Introduction

Let \mathbf{P} denote the class of all infinite words $x \in \mathbb{A}^\omega$ over a finite alphabet \mathbb{A} such that for every finite coloring $\varphi : \mathbb{A}^+ \rightarrow C$ there exist $c \in C$ and a factorization $x = V_0V_1V_2 \cdots$ with $\varphi(V_i) = c$ for all $i \geq 0$. Such a factorization is called φ -monochromatic. In [5] we conjectured:

Conjecture 1. *Let x be an infinite word. Then $x \in \mathbf{P}$ if and only if x is (purely) periodic.*

Various partial results in support of Conjecture 1 were obtained in [5, 6, 14]. Given $x \in \mathbb{A}^\omega$, it is natural to consider the binary coloring $\varphi : \mathbb{A}^+ \rightarrow \{0, 1\}$ defined by $\varphi(u) = 0$ if u is a prefix of x

Email addresses: aldo.deluca@unina.it (Aldo de Luca), lupastis@gmail.com (Luca Q. Zamboni)

and $\varphi(u) = 1$ otherwise. Then any φ -monochromatic factorization is nothing more than a prefixal factorization of x , i.e., a factorization of the form $x = U_0U_1U_2 \cdots$ where each U_i is a non-empty prefix of x . Thus a first necessary condition for a word x to belong to \mathbf{P} is that x admit a prefixal factorization. Not all infinite words admit such a factorization including for instance the class of square-free words and the class of Lyndon words [5].

Thus in the study of the Conjecture 1, one can restrict to the class of words \mathcal{P}_1 admitting a prefixal factorization. But in fact more is true. It is shown that if $x \in \mathcal{P}_1$, then x has only finitely many distinct unbordered prefixes and admits a unique factorization in terms of its unbordered prefixes. This allows us to associate with each $x \in \mathcal{P}_1$ a new infinite word $\delta(x)$ on an alphabet corresponding to the finite set of unbordered prefixes of x . In turn, the word $\delta(x)$ may or may not admit a prefixal factorization. In case $\delta(x) \notin \mathcal{P}_1$, then $\delta(x) \notin \mathbf{P}$ and from this one may deduce that x itself does not belong to \mathbf{P} . This is for instance the case of the famous Thue-Morse infinite word $t = t_0t_1t_2 \cdots \in \{0, 1\}^\omega$ where t_n is defined as the sum modulo 2 of the digits in the binary expansion of n ,

$$t = 011010011001011010010 \cdots$$

The origins of t go back to the beginning of the last century with the works of A. Thue [15, 16] in which he proves amongst other things that t is *overlap-free*, i.e., contains no word of the form uuu' where u' is a non-empty prefix of u . It is readily checked that t admits a prefixal factorization, in particular t may be factored uniquely as $t = V_0V_1V_2 \cdots$ where each $V_i \in \{0, 01, 011\}$. On the other hand as is shown later (see Example 4), the *derived word* $\delta(t)$ is the square-free ternary Thue-Morse word fixed by the morphism $1 \mapsto 123, 2 \mapsto 13, 3 \mapsto 1$. Hence $\delta(t) \notin \mathcal{P}_1$. This in turn implies that $t \notin \mathbf{P}$. Concretely, consider the coloring $\varphi' : \{0, 1\}^+ \rightarrow \{0, 1, 2\}$ defined by $\varphi'(u) = 0$ if u is a prefix of t ending with 0, $\varphi'(u) = 1$ if u is a prefix of t ending with 1, and $\varphi'(u) = 2$ otherwise. We claim that t does not admit a φ' -monochromatic factorization. In fact, suppose to the contrary that $t = V_0V_1V_2 \cdots$ is a φ' -monochromatic factorization. Since V_0 is a prefix of t , it follows that there exists $a \in \{0, 1\}$ such that each V_i is a prefix of t terminating with a . Pick $i \geq 1$ such that $|V_i| \leq |V_{i+1}|$. Then $aV_iV_i \in \text{Fact}(t)$. Writing $V_i = ua$, (with u empty or in $\{0, 1\}^+$), we see $aV_iV_i = auaua$ is an overlap, contradicting that t is overlap-free.

Thus, in the study of Conjecture 1, one can further restrict to the subset \mathcal{P}_2 of \mathcal{P}_1 consisting of all $x \in \mathcal{P}_1$ for which $\delta(x) \in \mathcal{P}_1$. In this case, one can define a second derived word $\delta^2(x) = \delta(\delta(x))$ which again may or may not belong to \mathcal{P}_1 . In case $\delta^2(x) \notin \mathcal{P}_1$, then not only is $\delta^2(x) \notin \mathbf{P}$, but as we shall see neither are $\delta(x)$ and x . Continuing in this way, we are led to consider the class \mathcal{P}_∞ of all words x in \mathcal{P}_1 such that $\delta^n(x) \in \mathcal{P}_1$ for all $n \geq 1$. We show that $\mathbf{P} \subset \mathcal{P}_\infty$, so in other words any potential counter-example to our conjecture is amongst the non-periodic words belonging to \mathcal{P}_∞ . However, $\mathbf{P} \neq \mathcal{P}_\infty$. In fact, we prove in Sect. 6 that a large class of Sturmian words (nonsingular Sturmian words) belong to \mathcal{P}_∞ , while as shown in [5], no Sturmian word belongs to \mathbf{P} .

The paper is organized as follows: In Sect. 2 we give a brief overview of some basic definitions and notions in combinatorics on words which are relevant to the subsequent material. In Sect. 3 we

study the basic properties of words admitting a prefixal factorization and in particular show each admits a unique factorization in terms of its finite set of unbordered prefixes. From this we define the derived word $\delta(x)$. We prove amongst other things that if x is a fixed point of a morphism, then the same is true of $\delta(x)$.

In Sect. 4 we recursively define a nested sequence $\cdots \subset \mathcal{P}_{n+1} \subset \mathcal{P}_n \subset \cdots \subset \mathcal{P}_1$ where $\mathcal{P}_{n+1} = \{x \in \mathcal{P}_n \mid \delta(x) \in \mathcal{P}_n\}$, and study some basic properties of the set $\mathcal{P}_\infty = \bigcap_{n \geq 1} \mathcal{P}_n$.

In Sect. 5 we study the connection between the class \mathbf{P} and the class \mathcal{P}_∞ and show that $\mathbf{P} \subset \mathcal{P}_\infty$. We also show that if $x \in \mathcal{P}_\infty$, then x is uniformly recurrent, from which we recover a result previously proved in [5] via different techniques.

Sect. 6 is devoted to prefixal factorizations of Sturmian words. Any Sturmian word $x \neq aS$, where $a \in \{0, 1\}$ and S a standard Sturmian word, admits a prefixal factorization. The main result of the section is that a Sturmian word x belongs to \mathcal{P}_∞ if and only if x is nonsingular, i.e., no proper suffix of x is a standard Sturmian word.

2. Notation and Preliminaries

Given a non-empty set \mathbb{A} , or *alphabet*, we let \mathbb{A}^* denote the set of all finite words $u = u_1 u_2 \cdots u_n$ with $u_i \in \mathbb{A}$. The quantity n is called the *length* of u and is denoted $|u|$. The *empty word*, denoted ε , is the unique element in \mathbb{A}^* with $|\varepsilon| = 0$. We set $\mathbb{A}^+ = \mathbb{A}^* \setminus \{\varepsilon\}$. For each word $v \in \mathbb{A}^+$, let $|u|_v$ denote the number of occurrences of v in u . In the following we suppose that the alphabet \mathbb{A} is finite even though several results hold true for any alphabet.

Let $u \in \mathbb{A}^*$. A word v is a *factor* of u if there exist words r and s such that $u = rvs$; v is a *proper factor* if $v \neq u$. If $r = \varepsilon$ (resp., $s = \varepsilon$), then v is called a *prefix* (resp., a *suffix*) of u .

Given words $u, v \in \mathbb{A}^+$ we say v is a *border* of u if v is both a proper prefix and a proper suffix of u . In case u admits a border, we say u is *bordered*. Otherwise u is called *unbordered*.

Let \mathbb{A}^ω denote the set of all one-sided infinite words $x = x_0 x_1 \cdots$ with $x_i \in \mathbb{A}$, $i \geq 0$.

Given $x \in \mathbb{A}^\omega$, let $\text{Fact}^+(x) = \{x_i x_{i+1} \cdots x_{i+j} \mid i, j \geq 0\}$ denote the set of all non-empty *factors* of x . Moreover, we set $\text{Fact}(x) = \{\varepsilon\} \cup \text{Fact}^+(x)$. The *factor complexity* of x is the map $\lambda_x : \mathbb{N} \rightarrow \mathbb{N}$ defined as follows: for any $n \geq 0$

$$\lambda_x(n) = \text{card}(\mathbb{A}^n \cap \text{Fact}(x)),$$

i.e., $\lambda_x(n)$ counts the number of distinct factors of x of length n . A factor u of a finite or infinite word x is called *right special* (resp., *left special*) if there exist two different letters a and b such that ua and ub (resp., au and bu) are factors of x . A factor u of x which is right and left special is called *bispecial*.

Given $x = x_0 x_1 x_2 \cdots \in \mathbb{A}^\omega$. A factor u of $x \in \mathbb{A}^\omega$ is called *recurrent* if u occurs in x an infinite number of times, and is called *uniformly recurrent* if there exists an integer k such that every factor of x of length k contains an occurrence of u . An infinite word x is called *recurrent* (resp., *uniformly recurrent*) if each of its factors is recurrent (resp., uniformly recurrent).

Let $x \in \mathbb{A}^\omega$ and \mathcal{S} denote the *shift operator*. The *shift orbit* of x is the set $\text{orb}(x) = \{\mathcal{S}^k(x) \mid k \geq 0\}$, i.e., the set of all suffixes of x . The *shift orbit closure* of x is the set $\Omega(x) = \{y \in \mathbb{A}^\omega \mid \text{Fact}(y) \subseteq \text{Fact}(x)\}$.

An infinite word x is called (purely) *periodic* if $x = u^\omega$ for some $u \in A^+$, and is called *ultimately periodic* if $x = vu^\omega$ for some $v \in \mathbb{A}^*$, and $u \in \mathbb{A}^+$. As is well known, an ultimately periodic word which is non-periodic is not recurrent. The word x is called *aperiodic* if x is not ultimately periodic.

We say that two finite or infinite words $x = x_0x_1\dots$ and $y = y_0y_1\dots$ on the alphabets \mathbb{A} and \mathbb{A}' respectively are *word isomorphic*, or simply *isomorphic*, and write $x \simeq y$, if there exists a bijection $\phi : \mathbb{A} \rightarrow \mathbb{A}'$ such that $y = \phi(x_0)\phi(x_1)\dots$.

For all definitions and notation not explicitly given in the paper, the reader is referred to the books [1, 12, 13].

3. Prefixal factorizations

Definition 1. We say that an infinite word $x \in \mathbb{A}^\omega$ admits a prefixal factorization if x has a factorization

$$x = U_0U_1U_2\dots$$

where each U_i , $i \geq 0$, is a non-empty prefix of x .

Some properties of words having a prefixal factorization have been proved in [5]. We mention in particular the following:

Lemma 2. Let $x \in \mathbb{A}^\omega$ be an infinite word having a prefixal factorization. Then the first letter of x is uniformly recurrent.

Given $x = x_0x_1x_2\dots \in \mathbb{A}^\omega$, we let $UP(x)$ denote the set of all (non-empty) unbordered prefixes of x .

Proposition 3. Let $x = x_0x_1x_2\dots \in \mathbb{A}^\omega$. The following conditions are equivalent:

1. x admits a prefixal factorization.
2. x admits a unique factorization of the form $x = U_0U_1U_2\dots$ with $U_i \in UP(x)$ for each $i \geq 0$.
3. $\text{card}(UP(x)) < +\infty$.

Proof. Let us first prove that if x admits a factorization $x = U_0U_1U_2\dots$ with $U_i \in UP(x)$, then such a factorization is necessarily unique. Indeed, suppose that there exists a different factorization $x = U'_0U'_1U'_2\dots$ with $U'_i \in UP(x)$. Let $n \geq 0$ be the first integer such that $U_n \neq U'_n$. Without loss of generality we suppose that $|U_n| > |U'_n|$. We can write $U_n = U'_nU'_{n+1}\dots U'_{n+p}\xi$, with $p \geq 0$ and ξ prefix of U'_{n+p+1} . Hence, U_n is bordered, a contradiction.

We will now show that $3. \Rightarrow 2. \Rightarrow 1. \Rightarrow 3.$

$3. \Rightarrow 2.$ We begin by assuming $\text{card}(UP(x)) < +\infty$ and show how to construct a factorization of x in terms of unbordered prefixes of x . We define recursively an infinite sequence $U_0, U_1, U_2, \dots \in UP(x)$ such that $U_0 U_1 \dots U_n$ is a prefix of x for each $n \geq 0$, U_0 is the longest unbordered prefix of x , and for $n \geq 1$, U_n is the longest unbordered prefix of x which is a prefix of $(U_0 \dots U_{n-1})^{-1}x$. For $n = 0$ we simply set U_0 to be the longest unbordered prefix of x . Note U_0 is well defined since $\text{card}(UP(x)) < +\infty$. For the inductive step, let $n \geq 0$ and suppose we have defined U_0, \dots, U_n with the required properties. We show how to construct U_{n+1} . Let V be the prefix of x of length $|U_0 \dots U_n| + 1$. Then since $|V| > |U_0|$ it follows that V is bordered. Let v denote the shortest border of V . Then $v \in UP(x)$ and by induction hypothesis that U_n is unbordered it follows that $|v| = 1$. In other words, $(U_0 \dots U_n)^{-1}x$ begins with an unbordered prefix of x . Thus we define U_{n+1} to be the longest unbordered prefix of x which is a prefix of $(U_0 \dots U_n)^{-1}x$. It follows immediately that $U_0 \dots U_n U_{n+1}$ is a prefix of x . Thus we have shown that $3. \Rightarrow 2.$

$2. \Rightarrow 1.$ This implication is trivially true.

$1. \Rightarrow 3.$ If $x = V_0 V_1 V_2 \dots$ is a prefixal factorization of x , then each prefix of x longer than $|V_0|$ is necessarily bordered. Hence, $\text{card}(UP(x)) \leq |V_0|$. \square

A direct proof of the equivalence of conditions 1. and 3. in the preceding proposition is in [5, Lemma 3.7]. We also observe that an infinite word having a finite number of unbordered factors is purely periodic [10].

Let \mathcal{P}_1 denote the set of all infinite words $x = x_0 x_1 x_2 \dots$ over any finite alphabet satisfying any one of the three equivalent conditions given in Proposition 3. For $x \in \mathcal{P}_1$ let

$$x = U_0 U_1 U_2 \dots \quad (1)$$

be the unique factorization of x with $U_i \in UP(x)$ for $i \geq 0$. Let $UP'(x) = \{U_i \mid i \geq 0\} \subseteq UP(x)$, and set $n_x = \text{card}(UP'(x))$.

Given distinct elements $U, V \in UP'(x)$, we write $U \prec V$ if

$$\min\{i \mid U_i = U\} < \min\{i \mid U_i = V\},$$

in other words if the first occurrence of U in (1) is before the first occurrence of V in (1). Let

$$\phi : \{1, 2, \dots, n_x\} \rightarrow UP'(x)$$

denote the unique order preserving bijection. We define $\delta(x) \in \{1, 2, \dots, n_x\}^\omega$ by

$$\delta(x) = \phi^{-1}(U_0)\phi^{-1}(U_1)\phi^{-1}(U_2) \dots$$

Clearly $\phi(\delta(x)) = x$. We call $\delta(x)$ the *derived word* of x with respect to the morphism induced by the bijection $\phi : \{1, 2, \dots, n_x\} \rightarrow UP'(x)$.

Example 1. Let $\mathbb{A} = \{0, 1\}$ and let f be the Fibonacci word over \mathbb{A} ,

$$f = 010010100100101001010010010100 \dots,$$

which is fixed by the morphism (Fibonacci morphism) defined by $0 \mapsto 01, 1 \mapsto 0$. It is readily verified that $UP(f) = UP'(f) = \{0, 01\}$ and that $01 \prec 0$. One has $n_f = 2$ and $\phi(1) = 01, \phi(2) = 0$. The unique factorization of f in terms of $UP(f)$ is

$$f = (01)(0)(01)(01)(0)(01)(0)(01)(01)(0)(01)(01)(0)(01)(01)(0) \dots$$

Hence,

$$\delta(f) = 1211212112112121121 \dots,$$

and $\delta(f) \simeq f$.

Example 2. Let $x = 121312112131212131211213121312112131212131211213 \dots$ denote the Tribonacci word fixed by the morphism defined by $1 \mapsto 12, 2 \mapsto 13, 3 \mapsto 1$. It is readily verified that $UP'(x) = UP(x) = \{1, 12, 1213\}$, and $1213 \prec 12 \prec 1$. It follows that $n_x = 3$ and $\phi(1) = 1213, \phi(2) = 12$, and $\phi(3) = 1$. The unique factorization of x in terms of $UP(x)$ begins with

$$x = (1213)(12)(1)(1213)(12)(1213)(12)(1)(1213)(1213)(12)(1)(1213)(12)(1213)(12)(1)(1213) \dots$$

and hence

$$\delta(x) = 123121231123121231 \dots.$$

Remark 1. In general, if $x \in \mathcal{P}_1$, the set $UP'(x)$ may be a proper subset of $UP(x)$. For instance, consider $x = 10f$ where f is the Fibonacci word. Then it is readily verified that $UP(x) = \{1, 10, 100\}$ while $UP'(x) = \{10, 100\}$.

We extend ϕ to a morphism $\phi : \{1, 2, \dots, n_x\}^+ \rightarrow UP'(x)^+$.

Lemma 4. *The morphism $\phi : \{1, 2, \dots, n_x\}^+ \rightarrow UP'(x)^+$ is injective.*

Proof. Suppose $w = \phi(v) = \phi(v')$ with $v, v' \in \{1, 2, \dots, n_x\}^+$. Then w factors as a product of elements in $UP'(x)$. Since any such factorization is necessarily unique, it follows that $v = v'$. \square

While, as is readily verified, every prefix w of x may be written uniquely as a product of unbordered prefixes of x , in general, as we saw in the example of $10f$ (see Remark 1), it may not be possible to factor w over $UP'(x)$. However, the following lemma shows that if w occurs in a prefixal factorization of x , then $w = \phi(v)$ for some factor v of $\delta(x)$.

Lemma 5. *Let $x = V_0V_1V_2 \dots$ be a prefixal factorization of x . Then there exists a (unique) factorization $\delta(x) = v_0v_1v_2 \dots$ such that $\phi(v_i) = V_i$ for each $i \geq 0$.*

Proof. Let $x = U_0U_1U_2\cdots$ be the factorization of x in unbordered prefixes. Define $r_0 = 0$ and $r_n = \sum_{i=0}^{n-1} |U_i|$. In other words, r_n corresponds to the position of U_n in the preceding factorization. Similarly we define $s_0 = 0$ and $s_n = \sum_{i=0}^{n-1} |V_i|$. Then we claim that $\{s_n \mid n \geq 0\} \subseteq \{r_n \mid n \geq 0\}$. In fact, suppose to the contrary that there exist indices i, j such that $r_j < s_i < r_{j+1}$. This implies that there exists $k \geq i$, such that a prefix of V_k (possibly all of V_k) is a proper suffix of U_j . This is a contradiction since U_j is unbordered. Thus we have shown that any prefixal factorization of x is also a factorization of x viewed as an infinite word over the alphabet $UP'(x)$, in other words. The result now follows. \square

Combining the two previous lemmas we obtain:

Corollary 6. *Let $x = V_0V_1V_2\cdots$ be a prefixal factorization of x . Then for each $i \geq 0$ there exists a unique factor v_i of $\delta(x)$ such that $\phi(v_i) = V_i$.*

As another consequence:

Corollary 7. *Suppose $x \in \mathcal{P}_1$ is a fixed point of a morphism $\tau : \mathbb{A}^+ \rightarrow \mathbb{A}^+$. Then there exists a morphism $\tau' : \{1, 2, \dots, n_x\}^+ \rightarrow \{1, 2, \dots, n_x\}^+$ fixing $\delta(x)$ such that $\phi \circ \tau' = \tau \circ \phi$.*

Proof. Applying τ to the unique factorization $x = U_0U_1U_2\cdots$ with $U_i \in UP'(x)$, we obtain a prefixal factorization $x = \tau(U_0)\tau(U_1)\tau(U_2)\cdots$. Writing $\delta(x) = a_0a_1a_2\cdots$ with $a_i \in \{1, 2, \dots, n_x\}$ and $\phi(a_i) = U_i$, by Lemma 5 there exists a unique factorization $\delta(x) = v_0v_1v_2\cdots$ such that $\phi(v_i) = \tau(U_i) = \tau(\phi(a_i))$ for each $i \geq 0$. The result now follows by defining $\tau'(a_i) = v_i$. \square

Example 3. As we saw in Example 2, the Tribonacci word x is in \mathcal{P}_1 . It follows from the previous corollary that $\delta(x)$ is also a fixed point of a morphism $\tau' : \{1, 2, 3\}^+ \rightarrow \{1, 2, 3\}^+$ which we can compute using the relation $\tau' = \phi^{-1} \circ \tau \circ \phi$ where τ denotes the Tribonacci morphism defined by $1 \mapsto 12, 2 \mapsto 13, 3 \mapsto 1$. So

$$\begin{aligned} 1 &\xrightarrow{\phi} 1213 \xrightarrow{\tau} 1213121 \xrightarrow{\phi^{-1}} 123 \\ 2 &\xrightarrow{\phi} 12 \xrightarrow{\tau} 1213 \xrightarrow{\phi^{-1}} 1 \\ 3 &\xrightarrow{\phi} 1 \xrightarrow{\tau} 12 \xrightarrow{\phi^{-1}} 2. \end{aligned}$$

Thus $\delta(x)$ is fixed by the morphism defined by $1 \mapsto 123, 2 \mapsto 1, 3 \mapsto 2$. It may be verified that in this example, $\delta(x)$ has the same factor complexity as x , namely it contains $2n + 1$ distinct factors of each length $n \geq 1$. However, unlike x which has a unique right and left special factor of each length, $\delta(x)$ has a unique left special factor of each length, and two right special factors of each length. It is also readily verified that $\delta(x) \in \mathcal{P}_1$. In fact $UP'(\delta(x)) = UP(\delta(x)) = \{1, 12, 123\}$, and $123 \prec 12 \prec 1$. Thus we obtain the infinite word $\delta^2(x) = \delta(\delta(x)) \in \{1, 2, 3\}^\omega$ which is a fixed point of a morphism τ'' verifying $\tau'' = \phi'^{-1} \circ \tau' \circ \phi'$. We compute τ'' as before:

$$1 \xrightarrow{\phi'} 123 \xrightarrow{\tau'} 12312 \xrightarrow{\phi'^{-1}} 12$$

$$\begin{aligned} 2 &\xrightarrow{\phi'} 12 \xrightarrow{\tau'} 1231 \xrightarrow{\phi'^{-1}} 13 \\ 3 &\xrightarrow{\phi'} 1 \xrightarrow{\tau'} 123 \xrightarrow{\phi'^{-1}} 1 \end{aligned}$$

and hence $\tau'' = \tau$, whence $\delta^2(x) = x$. Thus for each $n \geq 1$ we have that $\delta^n(x) \in \mathcal{P}_1$ and $\delta^n(x) = x$ for n even and $\delta^n(x) = \delta(x)$ for n odd.

Remark 2. We note that by Lemma 2, if $x \in \mathcal{P}_1$, the first letter x^F of x is uniformly recurrent in x , so that one can also define (see [8]) the bijection $\sigma : \{1, \dots, \text{card}(\mathcal{R}_{x^F})\} \rightarrow \mathcal{R}_{x^F}$, where \mathcal{R}_{x^F} is the finite set of the first returns of x^F to x^F in x , and define the derived word $D_{x^F}(x)$ with respect to the morphism induced by σ [5]. The two derived words $\delta(x)$ and $D_{x^F}(x)$ can be equal, as in the case of x equal to Fibonacci word; they can be different as in the case of Tribonacci word. In the case of a word aS where $a \in \{0, 1\}$ and S is a standard Sturmian word, one has that $\delta(aS)$ is not defined (cf. Lemma 19), whereas $D_a(aS)$ is defined.

4. A hierarchy of words admitting a prefixal factorization

We may recursively define a nested collection of words $\dots \subseteq \mathcal{P}_{n+1} \subseteq \mathcal{P}_n \subseteq \dots \subseteq \mathcal{P}_1$ by

$$\mathcal{P}_{n+1} = \{x \in \mathcal{P}_n \mid \delta(x) \in \mathcal{P}_n\}$$

and

$$\mathcal{P}_\infty = \bigcap_{n=1}^{\infty} \mathcal{P}_n.$$

Hence, a word $x \in \mathcal{P}_\infty$ if and only if $x \in \mathcal{P}_1$ and $\delta^n(x) \in \mathcal{P}_1$ for all $n \geq 1$. The previous example showed that the Tribonacci word belongs to \mathcal{P}_∞ . Similarly, following Example 1, the Fibonacci word also belongs to \mathcal{P}_∞ . In contrast, the following example shows that the Thue-Morse word does not belong to \mathcal{P}_∞ .

Example 4. The Thue-Morse word $t = 0110100110010110\dots$ is fixed by the morphism τ defined by $0 \mapsto 01, 1 \mapsto 10$. It is readily verified that $t \in \mathcal{P}_1$. In fact, $UP'(t) = UP(t) = \{0, 01, 011\}$ and $011 \prec 01 \prec 0$. Let $\phi : \{1, 2, 3\} \rightarrow UP'(t)$ be given by $1 \mapsto 011, 2 \mapsto 01, 3 \mapsto 0$. Then $\delta(t)$ is the fixed point of the morphism τ' which we compute as in Example 3:

$$\begin{aligned} 1 &\xrightarrow{\phi} 011 \xrightarrow{\tau} 011010 \xrightarrow{\phi^{-1}} 123 \\ 2 &\xrightarrow{\phi} 01 \xrightarrow{\tau} 0110 \xrightarrow{\phi^{-1}} 13 \\ 3 &\xrightarrow{\phi} 0 \xrightarrow{\tau} 01 \xrightarrow{\phi^{-1}} 2. \end{aligned}$$

We thus obtain that τ' is defined by $1 \mapsto 123, 2 \mapsto 13, 3 \mapsto 2$ which is the well known Hall morphism [9]. Thus $\delta(t)$ is the so-called *ternary Thue-Morse* word which is well known to be square-free (cf. [1], [12]). It follows (cf. [5]) that $\delta(t)$ does not admit a prefixal factorization, i.e., $\delta(t) \notin \mathcal{P}_1$. Thus $t \notin \mathcal{P}_2$ and hence $t \notin \mathcal{P}_\infty$.

Let \leq_p denote the prefixal order in \mathbb{A}^* , i.e., for $u, v \in \mathbb{A}^*$, $u \leq_p v$ if u is a prefix of v . We write $u <_p v$ if u is a proper prefix of v . For any word $u \in \mathbb{A}^+$ we let u^F denote the first letter of u .

Lemma 8. *Let $\Gamma = \{u_1, u_2, \dots, u_k\}$ be a finite set of unbordered words over the alphabet $\mathbb{A} = \{1, 2, \dots, k\}$ such that $u_k <_p u_{k-1} <_p \dots <_p u_1$. Let $u_1^F = 1$ and f be the morphism defined by $i \mapsto u_i$, $i = 1, \dots, k$. Then the word $x = f^\omega(1)$, fixed point of f , is such that $\delta(x) = x$.*

Proof. Let $x = f^\omega(1) = x_0 x_1 \dots x_n \dots$. Since x is fixed by f , one has

$$x = f(x) = f(x_0)f(x_1)\dots$$

Hence, x admits a unique factorization in unbordered prefixes of the set $\{f(x_i) \mid i \geq 0\} = UP'(x) \subseteq \Gamma$. Let $\phi : \{1, 2, \dots, n_x\} \rightarrow UP'(x)$ be the unique order preserving bijection. One has:

$$\delta(x) = \phi^{-1}(f(x_0))\phi^{-1}(f(x_1))\dots$$

Since $\phi^{-1}(f(x_n)) = x_n$, $n \geq 0$, it follows $\delta(x) = x$. □

Proposition 9. *For each $n \geq 1$ one has that \mathcal{P}_{n+1} is properly included in \mathcal{P}_n .*

Proof. Let t be the Thue-Morse word on two symbols $t = 0110100110010110\dots$. We have previously seen (see Example 4) that $t \in \mathcal{P}_1 \setminus \mathcal{P}_2$. Let F be the Fibonacci morphism and $F(t) = 01000100101000101001001\dots$. The word $F(t)$ has a prefixal factorization and $\delta(F(t)) \simeq t \in \mathcal{P}_1 \setminus \mathcal{P}_2$. It follows that $F(t) \in \mathcal{P}_2 \setminus \mathcal{P}_3$. It is easily verified that for any $n > 1$ one has that $\delta(F^n(t)) \simeq F^{n-1}(t)$. From this it easily follows that $\delta^n(F^n(t)) \simeq t$ and $\delta^{n+1}(F^n(t)) = \delta(t) \notin \mathcal{P}_1$. Hence, $F^n(t) \in \mathcal{P}_{n+1} \setminus \mathcal{P}_{n+2}$. □

Given an infinite word x with $\text{card}(UP(x)) < \infty$, we let $N(x)$ denote the length of the longest unbordered prefix of x . Now, for $x \in \mathcal{P}_\infty$, we define the map $\nu_x : \mathbb{N} \rightarrow \mathbb{N}$ by

$$\nu_x(n) = N(\delta^n(x)), \quad n \geq 0,$$

where $\delta^0(x) = x$. This is well defined and the sequence $(\nu_x(n))_{n \geq 0}$ is a sequence of natural numbers (≥ 2) which may be bounded or unbounded. If x is the Tribonacci word we have $(\nu_x(n))_{n \geq 0} = 4, 3, 4, 3, 4, 3, \dots$

Let $a \in \mathbb{A} = \{0, 1\}$ and put $b = 1 - a$. In the following for $a \in \{0, 1\}$, we let L_a be the injective endomorphism of $\{0, 1\}^*$ defined by

$$L_a : a \mapsto a, b \mapsto ab. \tag{2}$$

Proposition 10. *If $x \in \mathcal{P}_\infty$ and $(\nu_x(n))_{n \geq 0} = 2, 2, 2, 2, \dots$, then x is isomorphic to the Fibonacci word.*

Proof. Without loss of generality we can assume that x begins with 0. Since $N(x) = 2$, it follows that x begins with 01 and 01 is the longest unbordered prefix of x . It follows that x is a concatenation of 01 and 0 so we can write $x = L_0(x')$ for some binary word x' beginning with 1, where x' is isomorphic to $\delta(x)$. Since $N(\delta(x)) = N(x') = 2$, it follows that x' begins with 10 and 10 is the longest unbordered prefix of x' . Hence x' is a concatenation of 10 and 1 so that $x' = L_1(x'')$ where x'' begins with 0 and x'' is isomorphic to $\delta(x') = \delta^2(x)$. Since $N(x'') = 2$, it follows that x'' begins with 01 and 01 is the longest unbordered prefix of x'' . Thus $x'' = L_0(x''')$ for some x''' beginning with 1 and isomorphic to $\delta^3(x)$. Continuing in this way for each k we have $(L_0L_1)^k(0)$ is a prefix of x . Hence, x is isomorphic to Fibonacci word. \square

Let us observe that in general the sequence $(\nu(n))_{n \geq 0}$ does not determine x . For instance, let x be the word fixed by the morphism $0 \mapsto 00001, 1 \mapsto 0$ and y be the word fixed by the morphism $0 \mapsto 00101, 1 \mapsto 001$. Then $(\nu_x(n))_{n \geq 0} = (\nu_y(n))_{n \geq 0} = 5, 5, 5, \dots$

The following question naturally arises:

Question 1. What can be said about the nature of x if $(\nu_x(n))_{n \geq 0}$ is ultimately periodic? Is x necessarily a fixed point of a morphism? Conversely, if $x \in \mathcal{P}_\infty$ is a fixed point of a primitive morphism, is $(\nu_x(n))_{n \geq 0}$ ultimately periodic (in particular bounded) ?

5. A coloring problem

Let \mathbf{P} be the class of all infinite words x over any finite alphabet \mathbb{A} such that for every finite coloring $\varphi : \mathbb{A}^+ \rightarrow C$ there exists $c \in C$ and a factorization $x = V_0V_1V_2 \cdots$ with $\varphi(V_i) = c$ for all $i \geq 0$. Such a factorization is called φ -*monochromatic*. Thus if $x \notin \mathbf{P}$, then there exists a finite coloring $\varphi : \mathbb{A}^+ \rightarrow C$ such that for every factorization $x = V_0V_1V_2 \cdots$ we have $\varphi(V_i) \neq \varphi(V_j)$ for some $i \neq j$. Any such coloring will be called a *separating coloring* for x .

We conjectured [5] that \mathbf{P} coincides with the set of all periodic words. Partial results in this direction are given in [5] (see also [6, 14]).

Lemma 11. *Let $x \in \mathcal{P}_1$. The following holds:*

$$x \in \mathbf{P} \iff \delta(x) \in \mathbf{P}.$$

Proof. We begin by showing that if $\delta(x) \in \mathbf{P}$ then $x \in \mathbf{P}$. Since $x \in \mathcal{P}_1$, we have $x = \phi(\delta(x))$, where ϕ is the morphism induced by the unique order bijection $\phi : \{1, 2, \dots, n_x\} \rightarrow UP'(x)$. Thus if $x \notin \mathbf{P}$, then by the morphic invariance property [3, Proposition 4.1], one obtains that $\delta(x) \notin \mathbf{P}$.

We next prove the converse. Suppose $\delta(x) \notin \mathbf{P}$. Then there exists a separating coloring $\varphi : \mathbb{A}^+ \rightarrow C$ for $\delta(x)$. Put $C' = C \cup \{+\infty, -\infty\}$ where we assume $+\infty, -\infty \notin C$. The coloring φ induces a coloring $\varphi' : \mathbb{A}^+ \rightarrow C'$ defined as follows:

$$\varphi'(u) = \begin{cases} \varphi(\phi^{-1}(u)) & \text{if } u \text{ is a prefix of } x \text{ and } u = \phi(v) \text{ for some } v \in \text{Fact}(\delta(x)); \\ +\infty & \text{if } u \text{ is a prefix of } x \text{ and } u \notin \phi(\text{Fact}(\delta(x))); \\ -\infty & \text{if } u \text{ is not a prefix of } x. \end{cases}$$

We note that it follows from Lemma 4 that φ' is well defined. We now claim that φ' is a separating coloring for x . Suppose to the contrary that x admits a φ' -monochromatic factorization $x = V_0V_1V_2\cdots$ in non-empty factors. Since V_0 is a prefix of x , it follows that $\varphi'(V_0) \neq -\infty$, and hence $\varphi'(V_i) \neq -\infty$ for each $i \geq 0$. Thus the factorization $x = V_0V_1V_2\cdots$ is a prefixal factorization of x . It now follows from Lemma 5 that there exists a factorization $\delta(x) = v_0v_1v_2\cdots$ such that $\phi(v_i) = V_i$ for each $i \geq 0$. Since $\varphi'(V_i) = \varphi(v_i)$, it follows that $\delta(x)$ admits a φ -monochromatic factorization, a contradiction. \square

Theorem 12. *The following holds: $\mathbf{P} \subset \mathcal{P}_\infty$.*

Proof. Suppose that $x \notin \mathcal{P}_\infty$. Thus there exists some $n \geq 1$ such that $x \notin \mathcal{P}_n$. First suppose $x \notin \mathcal{P}_1$. Since x does not admit a prefixal factorization, one has $x \notin \mathbf{P}$ (see [5, Proposition 3.3]). Next suppose $x \in \mathcal{P}_1$ but $x \notin \mathcal{P}_n$ for some $n \geq 2$. Then $\delta^n(x) \notin \mathcal{P}_1$ and hence as above $\delta^n(x) \notin \mathbf{P}$. By an iterated application of the preceding lemma it follows that $x \notin \mathbf{P}$. \square

Let us observe that if x is a periodic word of \mathbb{A}^ω , then for every finite coloring x has a monochromatic factorization, so that by Theorem 12, or as one immediately verifies, $x \in \mathcal{P}_\infty$. From Theorem 12 one has that any counter-example to our Conjecture 1 belongs to the set \mathcal{P}_∞ , which is our main motivation for studying this class of words. However, the converse of Theorem 12 is not true. For instance, as proved in [5], Sturmian words do not belong to \mathbf{P} whereas, as we shall see in the next section (see Theorem 25), a large class of Sturmian words (nonsingular Sturmian words) belong to \mathcal{P}_∞ .

Lemma 13. *If $\varphi : \mathbb{A}^+ \rightarrow C$ is a separating coloring for x , then $\hat{\varphi} : \mathbb{A}^+ \rightarrow C \cup \{\infty\}$, with $\infty \notin C$, defined by*

$$\hat{\varphi}(u) = \begin{cases} \infty & \text{if } u \text{ is not a prefix of } x; \\ \varphi(u) & \text{if } u \text{ is a prefix of } x, \end{cases}$$

is a separating coloring for x .

Proof. Suppose that there exists a $\hat{\varphi}$ -monochromatic factorization

$$x = V_0V_1V_2\cdots$$

in non-empty factors V_i , $i \geq 0$. Since V_0 is a prefix of x , the preceding factorization has to be a φ -monochromatic prefixal factorization, a contradiction as φ is separating for x . \square

Proposition 14. *Let $x \in \mathbb{A}^\infty$ and $\Omega(x)$ the shift-orbit closure of x . If $x \notin \mathcal{P}_\infty$, there exists a separating coloring φ for x such that if $y \in \Omega(x)$ and $y \neq x$, then y has a φ -monochromatic factorization.*

Proof. By Theorem 12 one has $x \notin \mathbf{P}$. If $y \in \Omega(x)$ and $y \neq x$, then, as x is not periodic, y can always be factorized as $y = V_0 V_1 V_2 \cdots$ where each V_i , $i \geq 0$, is not a prefix of x . Let $\hat{\varphi}$ be the separating coloring for x defined in the preceding lemma. Then one has that $\hat{\varphi}(V_j) = \infty$ for all $j \geq 0$. \square

Proposition 15. *Let $x \in \mathbb{A}^\omega$ and let \mathcal{A} be any finite collection of words in $\mathcal{P}_\infty^c \cap \Omega(x)$, where \mathcal{P}_∞^c denotes the complement of \mathcal{P}_∞ . Then there exists a finite coloring $\varphi : \mathbb{A}^+ \rightarrow C$ such that for each $y \in \Omega(x)$, φ is a separating coloring for y if and only if $y \in \mathcal{A}$.*

Proof. The result is trivial if $\mathcal{A} = \emptyset$. Indeed in this case it is sufficient to consider the coloring $\varphi : \mathbb{A}^+ \rightarrow C$ defined as follows: for any $u \in \mathbb{A}^+$, $\varphi(u) = c \in C$. In this way any $y \in \Omega(x)$ will have a φ -monochromatic factorization. Let us then suppose that \mathcal{A} is not empty.

Let $\mathcal{A} = \{y_1, \dots, y_r\}$. Since for $1 \leq i < j \leq r$, $y_i \neq y_j$, there exists a positive integer k such that any word of \mathbb{A}^* of length $\geq k$ can be prefix of at most one of the words of \mathcal{A} . As $\mathcal{A} \in \mathcal{P}_\infty^c$ by Theorem 12, no word $y_i \in \mathcal{A}$ belongs to \mathbf{P} . Hence, for each $1 \leq i \leq r$ there exists a coloring $\varphi_i : \mathbb{A}^+ \rightarrow C_i$ which is separating for y_i . Let us observe that as $\mathcal{A} \subseteq \Omega(x)$ one has

$$\bigcup_{i=1}^r \text{Fact}(y_i) \subseteq \text{Fact}(x).$$

We can define a finite coloring φ on \mathbb{A}^+ as follows. For $u \in \mathbb{A}^+$,

$$\varphi(u) = \begin{cases} u & \text{if } |u| < k \text{ and } u \text{ is a prefix of at least one word of } \mathcal{A}; \\ \varphi_1(u) & \text{if } u \text{ is a prefix of } y_1 \text{ of length } \geq k; \\ \vdots & \vdots \\ \varphi_r(u) & \text{if } u \text{ is a prefix of } y_r \text{ of length } \geq k; \\ \infty & \text{if } u \text{ is not a prefix of any of the words of } \mathcal{A}. \end{cases}$$

Let us first prove that for each $y \in \mathcal{A}$, the coloring φ is separating. Let $y = y_i \in \mathcal{A}$ and suppose that there exists a φ -monochromatic factorization $y_i = V_0 V_1 V_2 \cdots$ where each V_i is non-empty. Since V_0 is a prefix of y_i , the preceding factorization has to be a prefixal factorization. If $\varphi(V_i) = \varphi(V_0)$ and $|V_0| < k$ it would follow that $y_i = V_0^\omega$ a contradiction because $y_i \notin \mathbf{P}$. Thus as V_0 is a prefix of y_i of length $\geq k$, it follows that $\varphi_i(V_j) = \varphi_i(V_0)$ for all $j \geq 0$ and this contradicts the fact that φ_i is separating for y_i .

Let us now prove that if $y \notin \mathcal{A}$, then y admits a φ -monochromatic factorization. Since for each $1 \leq i \leq r$, $y \neq y_i$ and y_i is not periodic, one easily derives that y can be factorized as $y = V_0 V_1 \cdots$, where each V_j , $j \geq 0$, is not prefix of any of the words $y_i \in \mathcal{A}$, $i = 1, \dots, r$. From this one has $\varphi(V_j) = \infty$ for all $j \geq 0$. \square

Question 2. Let $x \in \mathbb{A}^\omega$ be uniformly recurrent. Given a finite coloring $\varphi : \mathbb{A}^+ \rightarrow C$ does there exist a finite (possibly empty) set $\mathcal{A} \subset \Omega(x)$ such that for each $y \in \Omega(x)$ we have that y admits a φ -monochromatic factorization if and only if $y \notin \mathcal{A}$?

Let us observe that in the previous question the hypothesis that x is uniformly recurrent is necessary. Indeed, let x be word $x = 010^21^20^31^3 \dots$ and $\varphi : \{0, 1\}^+ \rightarrow \{0, 1, *\}$ the finite coloring defined for all $u \in \text{Fact}^+(x)$ by $\varphi(u) = 0$, if u begins with 0, $\varphi(u) = 1$ if u begins with 1, and if $u \notin \text{Fact}^+(x)$, by $\varphi(u) = *$. In the shift orbit closure $\Omega(x)$ of x there are infinitely many words $0^n 1^\omega$, $n \geq 1$ which do not admit a φ -monochromatic factorization.

Proposition 16. Let $x \in \mathcal{P}_\infty$. Then x is uniformly recurrent.

Proof. We show by induction on $n \geq 1$ that for any infinite word x over a finite alphabet if the prefix of x of length n is not uniformly recurrent, then $x \notin \mathcal{P}_\infty$. If the first letter of x is not uniformly recurrent in x , then by Lemma 2 the word x does not admit a prefixal factorization, hence $x \notin \mathcal{P}_1$.

Let $n \geq 1$, and suppose the result holds up to n . Let $x \in \mathbb{A}^\omega$ and suppose that the prefix u of x of length $n + 1$ is not uniformly recurrent in x . We will show that $x \notin \mathcal{P}_\infty$. If $x \notin \mathcal{P}_1$, we are done. So we may assume that $x \in \mathcal{P}_1$. Let $\phi : \{1, 2, \dots, n_x\} \rightarrow UP'(x)$ denote the unique order preserving bijection. Consider $\delta(x) \in \{1, 2, \dots, n_x\}^\omega$. Since $x = \phi(\delta(x))$ it follows that $\delta(x)$ is not uniformly recurrent. If v is a prefix of $\delta(x)$ which is uniformly recurrent in $\delta(x)$, then $\phi(v)$ is a uniformly recurrent prefix of x . Moreover, for every prefix v of $\delta(x)$ we have $|v| < |\phi(v)|$. Thus the shortest non-uniformly recurrent prefix of $\delta(x)$ is of length smaller than or equal to n . By induction hypothesis, $\delta(x) \notin \mathcal{P}_\infty$, and hence $x \notin \mathcal{P}_\infty$. \square

As a consequence of Proposition 16 and Theorem 12 we recover the following result first proved in [5]:

Corollary 17. If $x \in \mathcal{P}$ then x is uniformly recurrent.

6. The case of Sturmian words

A word $x \in \{0, 1\}^\omega$ is called *Sturmian* if it is aperiodic and *balanced*, i.e., for all factors u and v of x such that $|u| = |v|$ one has

$$||u|_a - |v|_a| \leq 1, \quad a \in \{0, 1\}.$$

Definition 18. Let $a \in \{0, 1\}$. We say that a Sturmian word is of type a if it contains the factor aa .

Clearly a Sturmian word is either of type 0 or of type 1, but not both.

Alternatively, a binary infinite word x is Sturmian if x has a unique left (or equivalently right) special factor of length n for each integer $n \geq 0$. In terms of factor complexity, this is equivalent

to saying that $\lambda_x(n) = n + 1$ for $n \geq 0$. As a consequence one derives that a Sturmian word x is *closed under reversal*, i.e., if u is a factor of x , then so is its reversal u^\sim (see, for instance, [13, Chap. 2]).

A Sturmian word x is called *standard* (or *characteristic*) if all its prefixes are left special factors of x . Since, as is well known, Sturmian words are uniformly recurrent, it follows that for any Sturmian word x there exists a standard Sturmian word S such that $\text{Fact}(x) = \text{Fact}(S)$.

Following [4] we say that a Sturmian word $x \in \{0, 1\}^\omega$ is *singular* if it contains a standard Sturmian word as a proper suffix, i.e., there exist $u \in \{0, 1\}^+$ and a standard Sturmian word S such that $x = uS$. It is readily verified that the previous factorization is unique. A Sturmian word which is not singular is said to be *nonsingular*.

Let $a \in \{0, 1\}$ and $b = 1 - a$. In the following for $a \in \{0, 1\}$, we consider the injective endomorphism L_a of $\{0, 1\}^*$ defined in (2) and the injective endomorphism R_a of $\{0, 1\}^*$ defined by

$$R_a : a \mapsto a, b \mapsto ba.$$

We recall [13, Chap. 2] that the monoid generated by L_a and R_a , with $a \in \{0, 1\}$ contains all endomorphisms f of $\{0, 1\}^*$ which preserve Sturmian words, i.e., the image $f(y)$ of any Sturmian word y is a Sturmian word.

Let us observe that for any infinite word x over $\{0, 1\}$ we have $L_a(x) = aR_a(x)$. If x is a Sturmian word and v is a left special factor of x , then $L_a(v)$ is a left special factor of $L_a(x)$. Thus if S is a standard Sturmian word, then so is $L_a(S) = aR_a(S)$. Conversely, if S is a standard Sturmian word of type a , then $S = L_a(S')$ for some standard Sturmian word S' .

Lemma 19. *A Sturmian word x belongs to \mathcal{P}_1 if and only if $x \neq aS$ where $a \in \{0, 1\}$ and S a standard Sturmian word.*

Proof. Indeed, as is well known (see [2, 11]) the unbordered factors of length greater than 1 of a Sturmian word x are of the form bUc with $\{b, c\} = \{0, 1\}$ and U a bispecial factor of x . Therefore, if $x = ax'$ has infinitely many unbordered prefixes, then x' begins with infinitely many bispecial factors of x , and hence x' is a standard Sturmian word. Conversely, let $x = aS$ where S is a standard Sturmian word and $a \in \{0, 1\}$. The Sturmian word aS does not admit a prefixal factorization. Indeed, aS begins with an infinite number of distinct prefixes of the form aUb with U bispecial and then a palindrome. It follows that aS begins with arbitrarily long unbordered prefixes and hence, by Proposition 3 does not admit a prefixal factorization. \square

We begin by reviewing some terminology which will be used in the proof of the following lemma. Let $x \in \mathbb{A}^\omega$ and $a \in \mathbb{A}$. A word $u \in \mathbb{A}^+$ is called a *left first return* to a in x if ua is a factor of x which begins and ends with a and $|ua|_a = 2$, i.e., the only occurrences of a in ua are as a prefix and as a suffix. A word $u \in \mathbb{A}^+$ is called a *right first return* to a in x if au is a factor of x which begins and ends with a and $|au|_a = 2$. A word $u \in \mathbb{A}^+$ is called a *complete return* to a in x if u is a factor of x which begins and ends with a and $|u|_a \geq 2$. It is a basic fact that if u is a

complete return to a in x , then ua^{-1} factors uniquely as a product of left first returns to a in x and that $a^{-1}u$ factors uniquely as a product of right first returns to a in x .

Lemma 20. *Let $x \in \{0, 1\}^\omega$ be a Sturmian word of type a with $\text{card}(UP(x)) < +\infty$, and $N(x)$ the length of the longest unbordered prefix of x . Then:*

- i) *If $N(x) = 2$, there exists a Sturmian word $y \in \{0, 1\}^\omega$ isomorphic to $\delta(x)$ and $x = L_a(y)$.*
- ii) *If $N(x) > 2$ and x begins with a , there exists a Sturmian word $y \in \{0, 1\}^\omega$ beginning with a such that $x = L_a(y)$, $N(y) < N(x)$, and $\delta(x) = \delta(y)$. Moreover L_a establishes a one-to-one correspondence between $UP(y)$ and $UP(x)$, i.e., $L_a : UP(y) \rightarrow UP(x)$ is a bijection.*
- iii) *If $N(x) > 2$ and x begins with b there exists a Sturmian word $y \in \{0, 1\}^\omega$ beginning with b such that $x = R_a(y)$, $N(y) < N(x)$, and $\delta(x) = \delta(y)$. Moreover R_a establishes a one-to-one correspondence between $UP(y)$ and $UP(x) \setminus \{b\}$, i.e., $R_a : UP(y) \rightarrow UP(x) \setminus \{b\}$ is a bijection.*

Proof. To prove i), suppose $N(x) = 2$. It follows that $UP(x) = \{c, cd\}$ where $\{c, d\} = \{0, 1\}$. Since x is of type a and admits a factorization over $UP(x)$ and hence contains the factor cc , it follows that $a = c$ and hence $UP(x) = \{a, ab\}$. Thus the unique factorization of x over $UP(x)$ is equal to the factorization of x according to left first returns to a which is well known to be Sturmian. Alternatively, there exists a unique Sturmian word y such that $x = L_a(y)$. It follows that $\delta(x)$ is word isomorphic to y .

To prove ii), suppose x begins with a and its longest unbordered prefix is of length $N(x) > 2$. Then x begins with aa ; in fact if x begins with ab , since x is of type a we would have that $UP(x) = \{a, ab\}$ contradicting our assumption that $N(x) > 2$. Having established that x begins with aa , it follows that there exists a unique Sturmian word y such that $x = L_a(y)$ and moreover y also begins with a . In fact, y is obtained from x by factoring x according to left first returns to a where one codes the left first return a by a , and the left first return ab by b .

Next we show that L_a establishes a bijection between $UP(y)$ and $UP(x)$. We use the key fact that if $u \in \{0, 1\}^+$ and ua is a factor of x which begins and ends with a , then there exists a unique factor v of y such that $u = L_a(v)$. In fact, ua is a complete return to a and hence v is obtained from u by factoring u as a product of left first returns to a . In particular, if $u \in \{0, 1\}^+$ is a factor of x which begins with a and ends with b , then $u = L_a(v)$ for some factor v of y .

We begin by showing that $L_a(UP(y)) \subseteq UP(x)$, i.e., that $L_a : UP(y) \rightarrow UP(x)$. So let u be an unbordered prefix of y . If $|u| = 1$, then $u = a$ and hence $L_a(u) = a$ which is an unbordered prefix of x . If $|u| > 1$, then u begins with a and ends with b , and hence $L_a(u)$ begins with a and ends with b . If $L_a(u)$ were bordered, then any border v of $L_a(u)$ would also begin with a and end with b , whence we can write $v = L_a(v')$ for some border v' of u , contradicting that u is unbordered. Since the mapping $L_a : UP(y) \rightarrow UP(x)$ is clearly injective, to show that it is a bijection it remains to show that the mapping is surjective. So assume u is a unbordered prefix of x ; we will show that $u = L_a(u')$ for some unbordered prefix of y . This is clear if $u = a$, in which

case $u' = a$. If $|u| > 1$, then u begins with a and ends with b and hence $u = L_a(u')$ for some prefix u' of y . Moreover, if u' were bordered (say v' is a border of u') then $L_a(v')$ is a border of u , a contradiction.

It follows that if u is the longest unbordered prefix of y , then $N(y) = |u| < |L_a(u)| \leq N(x)$ where the first inequality follows from the fact that u must contain an occurrence of both 0 and 1. Finally, since $\text{card}(UP(y)) < +\infty$, we have that y admits a factorization $y = U_0 U_1 U_2 \cdots$ with $U_i \in UP(y)$ which by definition is isomorphic to $\delta(y)$. Applying L_a we obtain

$$x = L_a(U_0) L_a(U_1) L_a(U_2) \cdots$$

with $L_a(U_i) \in UP(x)$ isomorphic to $\delta(x)$. Hence, $\delta(x) = \delta(y)$ as required. This completes the proof of ii).

Finally to prove iii), suppose x begins with b . Then there exists a unique Sturmian word y such that $x = R_a(y)$ and moreover y begins with b . As in the previous case, it is readily checked that $R_a : UP(y) \rightarrow UP(x) \setminus \{b\}$ is a bijection. The idea is that if $u \in \{0, 1\}^+$ and au is a factor of x which begins and ends with a , then there exists a unique factor v of y such that $u = R_a(v)$. In fact, au is a complete return to a and hence u factors uniquely as a product of right first returns to a . Thus in particular, if $u \in \{0, 1\}^+$ is a factor of x which begins with b and ends with a , then $u = R_a(v)$ for some factor v of y . Finally, as in the previous case we deduce that $N(y) < N(x)$ and $\delta(y) = \delta(x)$. \square

Remark 3. We note that if x is a Sturmian word with $\text{card}(UP(x)) < +\infty$ and $N(x) > 2$, then, applying repeatedly ii) and iii) of Lemma 20, we deduce that there exist a Sturmian word y and a morphism $f \in \{L_0, L_1, R_0, R_1\}^+$ such that $x = f(y)$, $N(y) = 2$, and $\delta(y) = \delta(x)$.

Corollary 21. *Let $x \in \{0, 1\}^\omega$ be a Sturmian word with $\text{card}(UP(x)) < +\infty$. Then $\delta(x)$ is again Sturmian.*

Proof. First suppose $N(x) = 2$. In this case the result follows immediately from i) of Lemma 20. Next suppose $N(x) > 2$. Then by Remark 3 there exists a Sturmian word y such that $N(y) = 2$ and $\delta(y) = \delta(x)$. Hence $\delta(x)$ is Sturmian. \square

Corollary 22. *If x is a standard Sturmian word, then so is $\delta(x)$.*

Proof. By Lemma 19 any standard Sturmian word x has a prefixal factorization. Thus $\text{card}(UP(x)) < \infty$. Moreover, if x is of type a , then it begins with the letter a . The proof is then easily obtained by making induction on $N(x)$. If $N(x) = 2$ by i) of Lemma 20 there exists a Sturmian word y isomorphic to $\delta(x)$ such that $x = L_a(y)$. Since x is a standard Sturmian word, it follows that y , as well $\delta(x)$, is a standard Sturmian word. Let us now suppose $N(x) > 2$. By ii) of Lemma 20 there exists a Sturmian word y such that $N(y) < N(x)$ and $x = L_a(y)$ and $\delta(y) = \delta(x)$. Since y is standard, by induction $\delta(y) = \delta(x)$ is a standard Sturmian word. \square

Remark 4. An infinite word x over the alphabet \mathbb{A} is called *episturmian* [7] if it is closed under reversal and x has at most one right special factor of each length. Corollary 21 cannot be extended to episturmian words. Indeed, as we observed in Example 3, in the case of Tribonacci word x , which is an episturmian word, $\delta(x)$ has a unique left special factor of each length and two right special factors of each length, so that $\delta(x)$ is not episturmian.

Lemma 23. *Let $x \in \{0, 1\}^\omega$ be a Sturmian word with $\text{card}(UP(x)) < +\infty$. Let $x = U_0U_1U_2 \cdots$ be the unique factorization of x over $UP(x)$. Then there exist distinct unbordered prefixes U and V of x with $|V| < |U|$ such that $\{U_i \mid i \geq 0\} = \{U, V\}$. Moreover U is the longest unbordered prefix of x and V is the longest proper unbordered prefix of U .*

Proof. Following Corollary 21 we have that $\delta(x)$ is a Sturmian word. In particular, writing $x = U_0U_1U_2 \cdots$ with $U_i \in UP(x)$, we have $\text{card}(\{U_i \mid i \geq 0\}) = 2$. Thus there exist distinct unbordered prefixes U and V of x with $|V| < |U|$ such that $\{U_i \mid i \geq 0\} = \{U, V\}$. Since U_0 is the longest unbordered prefix of x , it follows that $U_0 = U$. It remains to show that V is the longest proper unbordered prefix of U .

Without loss of generality we may assume that x is of type 0. We proceed by induction on the length $N(x)$ of the longest unbordered prefix of x to show that V is the longest proper unbordered prefix of U , where U and V are as above. If $N(x) = 2$, we have that $UP(x) = \{0, 01\}$ and the result follows taking $V = 0$ and $U = 01$. Next suppose $N(x) > 2$ and suppose that the result is true up to $N(x) - 1$. Then by ii) and iii) of Lemma 20 it follows that there exists a Sturmian word y such that $x = L_0(y)$ in case x begins with 0, and $x = R_0(y)$ in case x begins with 1. Moreover, again using ii) and iii) it follows that $N(y) < N(x)$. Hence, if U' denotes the longest unbordered prefix of y , and V' denotes the longest proper unbordered prefix of U' , then it follows by induction hypothesis that y factors over $\{U', V'\}$. If x begins with 0, then applying L_0 to this factorization we obtain a factorization of x over $\{L_0(U'), L_0(V')\}$ and

hence $U = L_0(U')$ and $V = L_0(V')$. Since $L_0 : UP(y) \rightarrow UP(x)$ is a bijection, we deduce that V is the longest proper unbordered prefix of U . Similarly, if x begins with 1, then applying R_0 to this factorization we obtain a factorization of x over $\{R_0(U'), R_0(V')\}$ and hence $U = R_0(U')$ and $V = R_0(V')$. Since $R_0 : UP(y) \rightarrow UP(x) \setminus \{1\}$ is a bijection and U begins with 1 and ends with 00, and the shortest prefix of U beginning with 1 and ending with 0 is a proper unbordered prefix of U , we deduce that V is the longest proper unbordered prefix of U . \square

Lemma 24. *Let $y \in \{0, 1\}^\omega$ be a Sturmian word, $f \in \{L_0, L_1, R_0, R_1\}^+$ and set $x = f(y)$. If x is singular, then y is singular. Conversely, if y is singular and of the form $y = u'S'$, where S' is a standard Sturmian word and $u' \in \{0, 1\}^+$ with $|u'| \geq 2$, then x is singular. More precisely, there exist a standard Sturmian word S and $u \in \{0, 1\}^+$ with $|u'| \leq |u|$ such that $x = uS$. Moreover, if f admits at least one occurrence of either L_0 or L_1 , then $|u'| < |u|$.*

Proof. Let us first suppose that S' is a standard Sturmian word and y is singular and of the form $y = u'S'$ with $u' \in \{0, 1\}^+$ and $|u'| \geq 2$. Since $|u'| \geq 2$, it follows that either $01S'$ or $10S'$ is a

suffix of y . In particular, u' must contain an occurrence of both 0 and 1. Whence for each $a \in \{0, 1\}$ we have that $|u'| < |L_a(u')|$ and $|u'| < |R_a(u')|$.

For $a \in \{0, 1\}$ let $g \in \{L_a, R_a\}$. Taking $g = L_a$ we have $g(y) = L_a(u'S') = L_a(u')L_a(S')$ and $|u'| < |L_a(u')|$. Taking $g = R_a$ we have $g(y) = R_a(u'S') = R_a(u')a^{-1}aR_a(S') = R_a(u')a^{-1}L_a(S')$ and $|u'| \leq |R_a(u')a^{-1}|$. Since $L_a(S')$ is a standard Sturmian word, one has that $g(y)$ is singular. Iterating we deduce that there exist a standard Sturmian word S and $u \in \{0, 1\}^+$ such that $x = f(y) = uS$ and $|u'| \leq |u|$. Moreover, if f admits at least one occurrence of either L_0 or L_1 , then $|u'| < |u|$.

Let us now suppose that $x = f(y)$ is singular, i.e., $x = uS$ with $u \in \{0, 1\}^+$ and S a standard Sturmian word. We wish to prove that y is singular. It suffices to prove the assertion for $f \in \{L_a, R_a\}$, $a \in \{0, 1\}$. Let us first take $f = L_a$. One has that x , as well as S , is a Sturmian word of type a beginning with the letter a . Hence the word u either ends with the letter b or ends with the letter a immediately followed by the letter a . Thus setting $S' = L_a^{-1}(S)$ and $u' = L_a^{-1}(u)$, one has $y = u'S'$. Since S' is a standard Sturmian word, one has that y is singular.

Let us now take $f = R_a$. One has that x , as well as S , is a Sturmian word of type a . We have to consider two cases. Case 1. The word u ends with the letter a . We can set $u' = R_a^{-1}(u)$. Since the first letter of S is a , we can write $y = u'aS'$ with $S = R_a(aS') = aR_a(S') = L_a(S')$. This implies that S' is a standard Sturmian word and that y is singular. Case 2. The word u ends with the letter b . Since S begins with the letter a , we can write $x = u_1baS''$ with $S = aS''$ and $u = u_1b$, where the word u_1 if it is different from ε , ends with the letter a . Setting $u' = R_a^{-1}(u_1)$ one has that $y = u'bS'$ where $bS = R_a(bS') = baR_a(S') = bL_a(S')$. Thus $S = L_a(S')$ and S' is a standard word. From this it follows that y is singular. \square

Remark 5. We note that the assumption in the preceding lemma that $|u'| \geq 2$ is actually necessary. For instance, if y is singular and of the form $y = aS'$ with $a \in \{0, 1\}$ and S' standard, then $R_a(y) = aR_a(S') = L_a(S')$ is nonsingular.

Theorem 25. *Let x be a Sturmian word. Then $x \in \mathcal{P}_\infty$ if and only if x is nonsingular.*

Proof. We begin by showing that if $x \notin \mathcal{P}_\infty$ then x is singular. For this we prove by induction on n , that if $x \notin \mathcal{P}_n$, then x is singular. For $n = 1$, we have that if $x \notin \mathcal{P}_1$, then by Lemma 19, $x = aS$ where $a \in \{0, 1\}$ and S is a standard Sturmian word. Thus x is singular. Next let $n \geq 2$, and suppose by inductive hypothesis that if y is a Sturmian word and $y \notin \mathcal{P}_{n-1}$, then y is singular. Let x be a Sturmian word with $x \notin \mathcal{P}_n$. By inductive hypothesis we can suppose $x \in \mathcal{P}_{n-1}$. In particular, $x \in \mathcal{P}_1$. If $N(x) = 2$, then by i) of Lemma 20 we have $x = L_a(y)$ where $a \in \{0, 1\}$ and y is Sturmian isomorphic to $\delta(x)$. Since $x \notin \mathcal{P}_n$ it follows that $\delta(x) \notin \mathcal{P}_{n-1}$, whence $y \notin \mathcal{P}_{n-1}$. Hence by induction hypothesis y is singular. Thus we can write $y = uS$ where $u \in \{0, 1\}^+$ and S a standard Sturmian word. Thus $x = L_a(y) = L_a(uS) = L_a(u)L_a(S)$ and since $L_a(S)$ is a standard Sturmian word, we deduce that x is singular as required.

If $N(x) > 2$, then, following Remark 3, there exist a Sturmian word y and a morphism $f \in \{L_0, L_1, R_0, R_1\}^+$ such that $x = f(y)$, $N(y) = 2$, and $\delta(y) = \delta(x)$. In particular, $\delta(y) \notin \mathcal{P}_{n-1}$

and hence $y \notin \mathcal{P}_n$. Thus applying i) of Lemma 20 as above, we deduce that y is singular. Hence there exist $u \in \{0, 1\}^+$ and a standard Sturmian word S such that $y = uS$. On the other hand, since $\delta(y)$ is defined (or equivalently $y \in \mathcal{P}_1$) we must have $|u| \geq 2$. Thus applying Lemma 24 we deduce that x is singular.

Conversely, suppose $x \in \{0, 1\}^\omega$ is a Sturmian word of the form $x = uS$ with $u \in \{0, 1\}^+$ and S a standard Sturmian word. We will show by induction on $|u|$ that $x \notin \mathcal{P}_\infty$. If $|u| = 1$, i.e., $x = aS$ for some $a \in \{0, 1\}$, then, by Lemma 19, $x \notin \mathcal{P}_1$, whence $x \notin \mathcal{P}_\infty$. Next let $n \geq 2$ and assume by induction hypothesis that if y is a Sturmian word of the form $y = u'S'$ where S' is a standard Sturmian word, and $u' \in \{0, 1\}^+$ with $|u'| < n$, then $y \notin \mathcal{P}_\infty$. Let x be a Sturmian word of the form $x = uS$ with S standard, $u \in \{0, 1\}^+$ and $|u| = n$. Since $n \geq 2$, by Lemma 19, x admits a prefixal factorization so that $\text{card}(UP(x)) < \infty$. We consider two cases. If $N(x) = 2$, then by i) of Lemma 20 we can write $x = L_a(y)$ where $a \in \{0, 1\}$ and y is Sturmian and isomorphic to $\delta(x)$. Since x is singular, it follows by Lemma 24 that y is singular. Thus we can write $y = u'S'$ for some $u' \in \{0, 1\}$ and some standard Sturmian word S' . If $|u'| = 1$, then $y \notin \mathcal{P}_1$, whence $\delta(x) \notin \mathcal{P}_1$ and hence $x \notin \mathcal{P}_\infty$. If $|u'| \geq 2$, again by Lemma 24 we deduce that $|u'| < |u|$ and hence by induction hypothesis we conclude that $y \notin \mathcal{P}_\infty$. Hence $\delta(x) \notin \mathcal{P}_\infty$ whence $x \notin \mathcal{P}_\infty$.

Finally suppose $N(x) > 2$. Then by Remark 3 there exist a Sturmian word y , and a morphism $f \in \{L_0, L_1, R_0, R_1\}^+$ such that $x = f(y)$ and $\delta(x) = \delta(y)$ and $N(y) = 2$. By i) of Lemma 20 there exists a Sturmian word y' isomorphic to $\delta(y)$ such that $y = L_a(y')$. Thus $x = f \circ L_a(y')$ and y' is isomorphic to $\delta(x)$. Since x is singular, by Lemma 24 we deduce that y' is singular. Thus we can write $y' = u'S'$ for some $u' \in \{0, 1\}^+$ and some standard Sturmian word S' . If $|u'| = 1$, then $y' \notin \mathcal{P}_1$, whence $\delta(x) \notin \mathcal{P}_1$ and hence $x \notin \mathcal{P}_\infty$. If $|u'| \geq 2$, then by Lemma 24 we deduce that $|u'| < n$, and hence by induction hypothesis we conclude that $y \notin \mathcal{P}_\infty$. Hence $\delta(x) \notin \mathcal{P}_\infty$ whence $x \notin \mathcal{P}_\infty$. \square

References

- [1] J.-P. Allouche, J. Shallit, Automatic Sequences: Theory, Applications, Generalizations, Cambridge University Press, Cambridge 2003
- [2] J. Berstel, A. de Luca, Sturmian words, Lyndon words and trees, *Theor. Comput. Sci.* **178** (1997) 171–203
- [3] T. C. Brown, Colorings of the factors of a word, preprint, Department of Mathematics, Simon Fraser University, Canada (2006).
- [4] M. Bucci, S. Puzynina, L.Q. Zamboni, Central sets generated by uniformly recurrent words, *Ergodic Theory Dynam. Systems* **35** (2015) 714–736

- [5] A. de Luca, E. Pribavkina, L.Q. Zamboni, A coloring problem for infinite words, *J. Combin. Theory (Ser. A)* **125** (2014) 306–332.
- [6] A. de Luca, L.Q. Zamboni, On some variations of coloring problems of infinite words, arXiv:1504.06807, April 2015
- [7] X. Droubay, J. Justin, G. Pirillo, Episturmian words and some constructions of de Luca and Rauzy, *Theor. Comput. Sci.* **255** (2001) 539–553
- [8] F. Durand, A characterization of substitutive sequences using return words, *Discrete Math.* **179** (1998) 89–101
- [9] M. Hall Jr., Generators and relations in groups. The Burnside problem, Lectures on Modern Mathematics, Vol. II, Wiley, New York, NY, 1964, pp. 42–92
- [10] T. Harju, M. Huova, L.Q. Zamboni, On generating binary words palindromically, *J. Combin. Theory (Ser. A)* **129** (2015) 142–159
- [11] T. Harju, D. Nowotka, Minimal Duval extensions, *Internat. J. Found. Comput. Sci.* **15** (2004) 349–354
- [12] M. Lothaire, Combinatorics on Words, Addison-Wesley, Reading MA, 1983. Reprinted by Cambridge University Press, Cambridge UK, 1997
- [13] M. Lothaire, Algebraic Combinatorics on Words, Encyclopedia of Mathematics and its Applications, vol. 90, Cambridge University Press, Cambridge UK, 2002
- [14] V. Salo, I. Törmä, Factor colorings of linearly recurrent words, ArXiv: 1504.0582, April 2015.
- [15] A. Thue, Über unendliche Zeichenreihen, *Norske Vid. Selsk. Skr. I Math-Nat. Kl.* **7** (1906) 1–22.
- [16] A. Thue, Über die gegenseitige Lage gleicher Teile gewisser Zeichenreihen, *Norske Vid. Selsk. Skr. I. Mat-Nat. Kl.* **1** (1912) 1–67
- [17] L. Q. Zamboni, A Note on Coloring Factors of Words, in Oberwolfach Report 37/2010, Mini-workshop: Combinatorics on Words, August 22-27, 2010, pp. 42–44